DOCUMENT RESUME

ED 481 664 TM 035 360

AUTHOR	Popp, Sharon E. Osborn; Ryan, Joseph M.
TITLE	The Effect of Benchmark Selection on the Assessed Quality of Writing.
PUB DATE	2002-10-00
NOTE	18p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 21-25, 2003).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	*Benchmarking; *Elementary School Students; Primary Education; *Scoring Rubrics; Selection; Writing Achievement; *Writing Evaluation; Writing Tests
IDENTIFIERS	*Writing Samples

ABSTRACT

The purpose of this study was to investigate whether the selection of benchmark writing samples influences the assessment of students' writing quality. More than 300 grade 3 writing samples were scored in two separate rating sessions. Within each scoring session, raters used a different set of benchmark writing samples. Raw ratings were analyzed using multifacet Rasch models. Raw ratings and Rasch parameter estimates were examined and compared for the two sets of ratings. Ratings were also compared to hypothetical performance standards to illustrate the impact of differential benchmark selection. The same writing samples received very different ratings when different benchmark papers were used in scoring despite the uniform rubric. Results imply that assessed quality of writing may depend more on the benchmarks chosen to define the rubric than on the rubric itself. Results confirm the need for continued investigation into sources of construct-irrelevant variance in the design and development of writing assessments and suggest caution in the use and interpretation of large-scale writing assessment scores. (Contains 3 tables, 3 figures, and 12 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 481 664

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

S.O. Popp

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

THE EFFECT OF BENCHMARK SELECTION ON THE ASSESSED QUALITY OF WRITING

Sharon E. Osborn Popp, Arizona State University
Joseph M. Ryan, Arizona State University West

Paper presented at a distinguished papers session of the
State and Regional Educational Research Associations at the annual meeting of the
American Educational Research Association
Chicago, Illinois, April, 2003

First presented at the annual meeting of the Arizona Educational Research Organization
Tempe, Arizona, October, 2002

TM035360

BEST COPY AVAILABLE

Abstract

The purpose of this study was to investigate whether the selection of benchmark writing samples influences the assessment of students' writing quality. Grade 3 writing samples were scored in two separate rating sessions. Within each scoring condition, raters used a different set of benchmark writing samples. Raw ratings were analyzed using multi-facet Rasch models. Raw ratings and Rasch parameter estimates were examined and compared for the two sets of ratings. Ratings were also compared to hypothetical performance standards to illustrate the impact of differential benchmark selection. The same writing samples received very different ratings when different benchmark papers were used in scoring, despite the uniform rubric. Results imply that assessed quality of writing may depend more on the benchmarks chosen to define the rubric, than on the rubric itself. Results confirm the need for continued investigation into sources of construct-irrelevant variance in the design and development of writing assessments and suggest caution in the use and interpretation of large-scale writing assessment scores.

The Effect of Benchmark Selection on the Assessed Quality of Student Writing

Introduction

Direct assessments of writing performance are increasingly included in large-scale testing programs, often with high-stakes consequences, despite concerns regarding reliability and validity (Gordon, Engelhard, Gabrielson, and Bernknopf, 1996; Mehrens, 1992). The purpose of this study was to investigate the role of benchmark writing samples in a direct assessment of writing. Benchmarks, also known as anchor papers, exemplars, or range-finders, are the writing samples chosen to define levels of performance in the scoring rubric. The chosen benchmarks operationalize the concepts described in the language of the scoring rubric. They define the standards of performance for a given assessment and serve as the rubric's surrogate reference points, against which all samples are judged

The consistent application of the scoring rubric is considered essential to the validity and meaningful interpretation of scores for performance assessments (see e.g., Brennan and Johnson, 1995; Messick, 1995). The particular benchmarks chosen to represent levels of performance in the rubric would appear to be highly related to score outcome. However, research regarding the role of benchmarks in scoring direct writing assessments is surprisingly limited. We sought to investigate whether, and to what extent, benchmarks influence the ratings of students' writing. In this study, the same Grade 3 writing samples were scored in two separate rating sessions. Within each scoring condition, raters used a different set of benchmark writing samples. The two sets of benchmarks represented the same rubric but one set of benchmarks was chosen from

within the set of all Grade 3 papers and the second set was selected from a set of cross-grade papers, containing random subsets of papers from Grades 3, 5, and 8.

Method

Design

Grade 3 students produced writing samples in response to a Narrative mode prompt for a district-wide assessment of writing performance. Benchmarks were chosen from the set of Grade 3 Narrative writing samples and were used as the Within-grade benchmarks in scoring all Grade 3 papers. In addition, classrooms of Grade 5 and 8 students were randomly selected to respond to the same Narrative writing prompt as the Grade 3 students. The Narrative writing samples from Grade 5 and 8 students were combined with a random subset of Grade 3 student samples. Benchmark writing samples were chosen from this combined set of Grades 3, 5, and 8 papers and used in scoring this Across-grades set of papers.

Over 300 Grade 3 writing samples had two sets of ratings: one set scored against the Within-grade Benchmarks, and one set scored against the Across-grades Benchmarks. Raw ratings, as well as ability (theta) parameter estimates obtained in multi-facet Rasch model analyses, were compared between the two sets of ratings.

Data

All subjects that produced writing samples used in this study were Grades 3, 5, and 8 students from a large metropolitan school district. The writing assessments were part of an on-going district-wide assessment program intended to reflect progress toward curricular objectives in writing and language arts. For each grade, a writing prompt from a different discourse mode was presented. All Grade 3 students responded to a Narrative

mode prompt. Randomly selected classrooms of Grade 5 and Grade 8 students also responded to the same Narrative mode prompt as the Grade 3 students. The Narrative mode prompt is shown in Figure 1. There were 317 Grade 3 samples that were rated with the Within-grade benchmarks and again with the Across-Grades benchmarks. The Across-grades set of ratings also included 180 Grade 5 Narrative writing samples and 172 Grade 8 Narrative writing samples.

Think of something you have done, a special place you have been, or a special person you have known that has created a memory for you. Describe your feelings and why it was important to you.

Figure 1. Narrative mode writing prompt.

Students responded to the writing prompts in December, 1998. Assessments were administered over two, separate 50 minute periods. Randomly selected classrooms of Grade 5 and Grade 8 students were assessed over an additional two, separate 50 minute periods to obtain writing samples from the Narrative discourse mode. Teachers were required to read aloud the instructions as they appeared in a prepared teacher's manual.

Student writing samples were scored by professional raters from a commercial testing company in the early months of 1999. In each scoring session, two raters read and scored each paper. For any pair of score points that differed by more than 1 point, another rater was called upon to score the paper and provide a third rating. For this study, cases that required a third rating were excluded from analysis.

Raters scored the writing samples, using a six-point, six- trait rubric (Spandel, 1996). The six writing traits evaluated were:

1. Ideas (well-developed, clear, and complete),
2. Organization (logical order, clear introduction and ending, effective transitions),
3. Voice (commitment to topic, originality, appropriate feeling and tone),
4. Word Choice (adds interest and understanding, enhances detail),
5. Sentence Fluency (sentences flow, have varied lengths, and ease reading), and
6. Conventions (minimal errors in grammar, punctuation, spelling, and format).

Benchmark papers were chosen to guide scoring for each separate grade level. Benchmarks were also chosen from the combined Grades 3, 5, and 8 Across-grades set of writing samples. Professional raters chose the benchmarks and the final choices were reviewed and approved by school district staff. Each of the six score points, for each analytic trait, was represented by a benchmark paper chosen from the set of writing samples to be evaluated.

Procedure

Raw ratings were analyzed using multi-facet Rasch models. Raw ratings and Rasch-estimated student abilities, trait difficulties, and rater leniency-severity parameters were examined. The multi-facet Rasch model is an extension of the Rasch model (Rasch, 1960/1980; Wright and Stone, 1979) that accommodates multiple facets in the analysis. Student ability is estimated while accounting for rater severity and analytic-trait difficulty. The multi-facet (also called many-facet and many-faceted) Rasch model (Linacre, 1989) is an extension of Rasch ordered-category and partial credit models (Andrich, 1978;

Masters, 1982; Wright and Masters, 1982) and its use has been demonstrated previously in analyzing assessments of writing (e.g., Engelhard, 1992). The multi-facet Rasch model that was employed in this study can be expressed as Equation 1,

$$\log(P_{nijk} / P_{nijk-1}) = B_n - R_i - T_j - F_k , \quad (1)$$

where P_{nijk} is equal to the probability of student n being rated k on trait j by rater i, P_{nijk-1} is equal to the probability of student n being rated k – 1 on trait j by rater i, B_n is the writing ability of student n, R_i is the severity of rater i, T_j is the difficulty of analytic trait j, and F_k is the difficulty of rating threshold k, relative to rating threshold k – 1. Observed ratings are transformed into a linear logistic scale (in log-odds units, or logits) that ranges from $-\infty$ to $+\infty$. Perfect scores and zero scores are eliminated from analysis because they are non-estimable. Estimated student abilities, rater severity, and trait difficulty can be located along this scale and compared to each other. The distributions of latent trait locations within each ratings set for students, raters, and traits were examined.

The ratings from the Within-grade scoring were compared to the ratings from the Grade 3 subset of the Across-grades scoring. Raw ratings and Rasch parameter estimates were examined and compared between the different benchmark paper conditions. Rasch student-ability locations from each benchmark condition analysis were compared using a *t*-test for dependent samples. Patterns among the rater severities and trait difficulties within each benchmark conditions were examined, as well.

To illustrate the impact of benchmark selection on the assessed quality of student writing, the ratings sets are compared against hypothetical performance standards.

Contingency tables are provided to show the classifications of students (i.e., At or above standard or Below standard) based on two sets of ratings for the same papers. The proportion of misclassified students is reported for each of two hypothetical performance standards.

Results

Ratings of the same essays differed in magnitude and relative rank when scored against different sets of benchmarks. Not surprisingly, raw ratings were higher for the papers rated against the Within-grade benchmarks, with a mean of summed score-points of 20.7 (SD = 3.76), compared to 17.0 (SD = 4.32) for the same papers rated against the Across-grades benchmarks. The correlation between raw scores was .763 (r^2 = .5825). Rasch student-ability location estimates were also significantly higher (M = -2.57, SD = 3.88) for the Within-grade benchmark condition than the Across-grades benchmark condition (M = -3.84, SD = -3.52), with a t (df = 316) of 8.769, $p < .001$, $\alpha = .05$. As with the raw scores, the rank-ordering of student-ability locations differed between the Grade 3 Within-grade and Across-grades benchmark conditions, with a correlation between estimates of .762 (r^2 = .5806).

The distributions of rater-severity parameter estimates, or locations along a leniency-severity continuum (expressed in logits), were not remarkably different between the two benchmark conditions. Most raters in each set differed significantly from each other, with significant fixed chi-square values for the rater facet in both analyses, $\chi^2(5) = 236.2$, $p < .01$, $N = 6$ for the Within-grade condition and $\chi^2(11) = 892.5$, $p < 01$, $N = 12$ for the Across-grades condition. None of the six rater-severity locations had outfit mean-square statistics indicating misfit (≥ 3.00) in the Within-grade analysis. Only one rater-

severity location (A9) had a high standardized outfit mean-square value in the Narrative analysis (standardized outfit of 4.00, outfit mean-square of 1.3). Re-analysis after removal of this rater failed to reach convergence, so results are reported on the original analysis. Rater-severity locations, intentionally centered at zero, spanned less range for the 6 raters in the Within-grade benchmark condition ($\underline{M} = 0$; $\underline{SD} = 0.708$), than for the 12 raters in the Across-grades benchmark condition ($\underline{M} = 0$; $\underline{SD} = 0.786$).

The relative difficulty of the six analytic traits also differed considerably, depending on whether the samples were scored against the Within-grade benchmarks or the Across-grades benchmarks. Figure 2 shows the trait-difficulty locations (intentionally centered at zero in both analyses) estimated for the Within-grade and Across-grades ratings sets. Also, the range of difficulty is more restricted for the Within-grade benchmark type, with location estimates ranging from -1.01, for the least difficult trait, Word Choice, to +.96, for the most challenging trait of Conventions. The range of difficulty for the Across-grades benchmark type extends from -1.82, for Voice, to +2.36, for Conventions. Consequently, the trait locations, intentionally centered at zero in both analyses, were more widely dispersed ($\underline{M} = 0$; $\underline{SD} = 1.4228$) under the Across-grades condition than the Within-grade condition ($\underline{M} = 0$; $\underline{SD} = 0.6789$). Trait-difficulty locations under the different benchmark conditions are most different for Voice and Conventions(with differences of 1.42 and -1.40, respectively). Table 1 shows the trait-difficulty locations, along with their differences (Within - Across). For each benchmark condition, most analytic traits differed significantly among themselves, with significant fixed chi-square values for the trait facet in both analyses, $\chi^2(5) = 244.7$, $\underline{p} < .01$ for the Within-grade condition and $\chi^2(5) = 2455.9$, $\underline{p} < 01$ for the Across-grades condition.

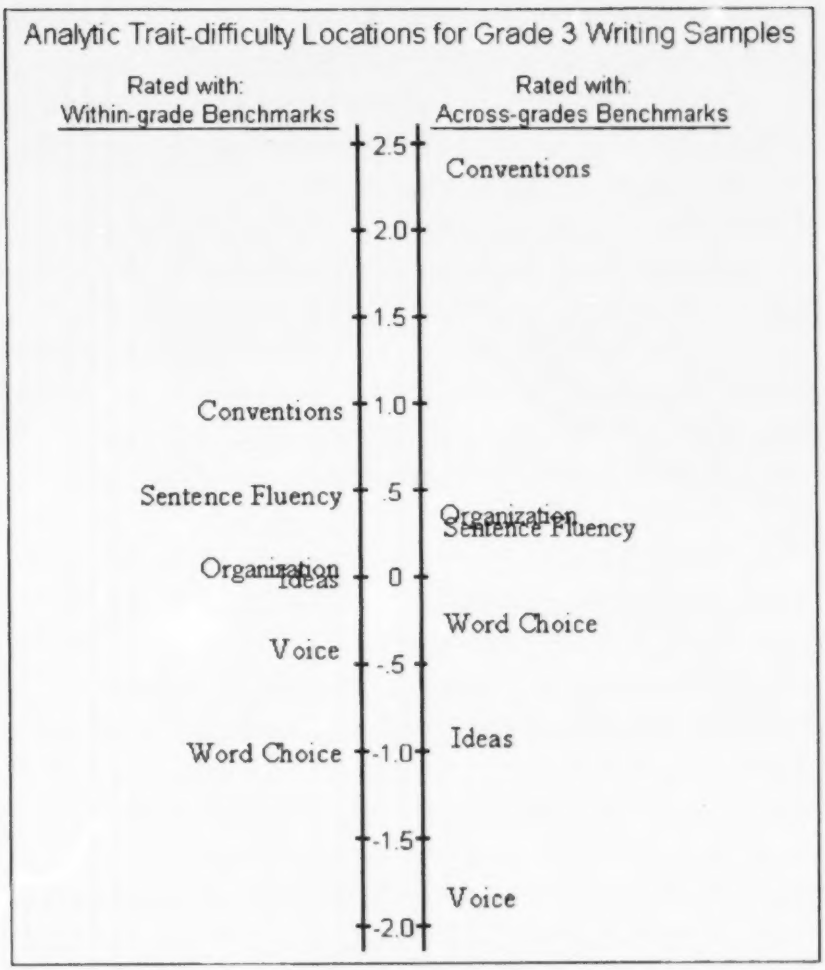

Figure 2. Trait-difficulty locations for Within-grade and Across-grades ratings sets for Grade 3 writing samples.

Table 1

Grade 3: Trait-difficulty Locations for each Analytic Trait by Benchmark Type

	Benchmark Condition		
Analytic Trait	Within-grade (SE)	Across-grades (SE)	Within - Across
Ideas	.00 (.10)	-.95 (.07)	.95
Organization	.01 (.10)	.38 (.06)	-.37
Voice	-.40 (.10)	-1.82 (.06)	1.42
Word Choice	-1.01 (.10)	-.30 (.07)	-.71
Sentence Fluency	.45 (.10)	.32 (.06)	.13
Conventions	.96 (.09)	2.36 (.06)	-1.40
Mean	0	0	0
Standard Deviation	1.42	.68	1.05

The largest shifts in difficulty on traits between the two conditions were Conventions (higher difficulty for Grade 3 in Across-grades analysis), Voice, and Ideas (both lower in difficulty in Across-grades analysis). However, looking at the relationship between the estimated trait difficulties reveals that there is a moderately high relationship between the trait-difficulties, and a near-perfect relationship between them, if Word Choice and Organization are omitted. Figure 3 is a scatterplot of the trait-difficulty locations from the two benchmark condition ratings.

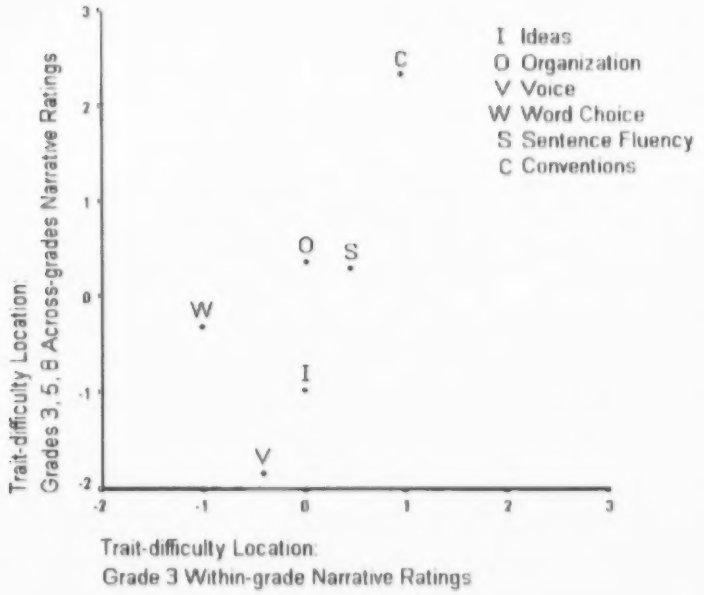

Figure 3. Trait-difficulty locations estimated for the Grade 3 Within-grade ratings and for the Grades 3, 5, and 8 Across-grades ratings, marked by analytic trait.

Given a compensatory standard set at an average raw score-point rating of 4 across all six analytic traits, fourteen percent of Grade 3 students would obtain inconsistent results (i.e., at or above standard on one mode and below standard on the other) on papers rated against different benchmarks. If a lower hypothetical standard is explored, such as an average raw score-point rating of 3 across analytic traits, 36% of students are classified differently between the two benchmark conditions. Most of the misclassification occurs with students who would be considered at or above the standard when rated against the Within-grade benchmarks. Under the higher standard, seventy-five percent of these students would be considered below standard when rated against Across-

grade benchmarks. Tables 2 and 3 display the contingencies for each hypothetical standard scenario, given the Grade 3 raw scores in this sample.

Table 2

Grade 3: Number of Students Meeting Hypothetical Compensatory Standard of Average Raw Score-point Rating of "4"when Scored Against Different Benchmark Papers

Classification		Across-grades Benchmarks		Total
		At or above Standard	Below Standard	
Within-grade Benchmarks	At or above standard	14	41	55
	Below Standard	3	259	262
Total		17	300	317

Table 3

Grade 3: Number of Students Meeting Hypothetical Compensatory Standard of Average Raw Score-point Rating of "3" when Scored Against Different Benchmark Papers

Classification		Across-grades Benchmarks		Total
		At or above Standard	Below Standard	
Within-grade Benchmarks	At or above standard	148	112	260
	Below Standard	3	54	57
Total		151	166	317

Discussion

The same writing samples, judged against the same rubric, received different ratings when different benchmark papers were used in scoring. The selection of different scoring benchmarks from either within or across grade levels did affect the assessment of student writing quality for the Grade 3 students in this study. Despite being scored against the same six-trait, six-point analytic rubric, Grade 3 Narrative writing samples received higher grades when scored against benchmark papers chosen from Grade 3 samples, than when scored against benchmark papers chosen from a combined set of samples from Grades 3, 5, and 8. If results on the two sets of ratings in this writing assessment were to be compared to a hypothetical standard, there would be a considerable difference in perceived success, depending on the benchmarks used for scoring.

The findings raise questions about the meaning and intentions underlying the rubric. The benchmarks chosen to represent the score-points in the rubric clearly reflected different interpretations, given the collection of writing samples to be scored. We might expect the writing samples of Grade 3 students to be rated lower when compared to the performance of Grade 5 or 8 students, than when compared to the writing of same-grade peers. However, we do not expect the same writing samples of Grade 3 students, scored against the same rubric, to be rated differently. Does the rubric reflect a broad construct of writing, representing all stages of writing ability, that spans the levels of performance that extend from novice, emerging writers to expert, accomplished writers? Or is the rubric intended to be interpreted at varying grade levels to reflect several narrow constructs that measure writing ability relative to grade-level expectations and curricular targets? In this study, the benchmarks translated the language of the rubric into two different

assessments; one that measured writing at grade level and one that measured a broader construct of writing ability. Student ratings differed substantially in magnitude and rank, and analytic traits differed in relative difficulty. The benchmarks operationalized the language of the rubric into two different assessments that reflected different contexts and perceptions of the construct of writing ability measured.

The relationship between the trait-difficulty locations also reveals that while the estimated difficulty of analytic traits differed considerably between the two scoring conditions, there are some traits may be related more strongly between the conditions than others. Word Choice, and to some degree, Organization, appeared to be possible outliers with respect to a potentially strong relationship between the trait estimates for the two benchmark conditions. This suggests that some aspects of a broader construct of writing ability may be comparable across grade levels, while other aspects of writing ability may be defined and assessed very differently depending on the writer's grade level. The results suggest a need for further research regarding the perceptions of raters with respect to rubric interpretation and the construct of writing.

The use of uniform criteria in writing scoring rubrics clearly does not ensure consistent application of the rubric. The standards of writing performance defined in the writing rubric imply a standards-based assessment framework. Benchmarks operationalize the rubric in the actual scoring of writing and are selected from the set of performances to be rated. The selection of benchmarks from a given set of examinee performances would imply a relative assessment framework. Results suggest that benchmark selection does transform the standards-based assessment framework defined by the writing rubric into a relative assessment framework.

Results of this study demonstrate that diversely defined ranges of least to highest quality could each be mapped to the generic language of a rubric. The selection of the benchmarks is an instrumental part of scoring and have a critical impact on scoring outcomes. Further research regarding the selection and use of benchmarks in scoring is needed to better understand the role of the benchmark as a critical element in direct writing assessment. Results confirm the need for continued investigation into sources of construct-irrelevant variance in the design and development of writing assessments and caution in the use and interpretation of large-scale writing assessment scores.

References

Andrich, D. (1978). A rating formulation for ordered categories. *Psychometrika, 43*, 357-374.

Brennan, R. L., & Johnson, E. G. (1995). Generalizability of performance assessments. Educational Measurement: Issues and Practice, 14 (4), 9-12.

Engelhard, G., Jr. (1992). The measurement of writing ability with a many-faceted Rasch model. Applied Measurement in Education, 5 (3), 171-191.

Gordon, B., Engelhard, G., Jr., Gabrielson, S., & Bernknopf, S. (1996). Conceptual issues in equating performance assessments: lessons from writing assessment. Journal of Research and Development in Education, 29 (2), 81-88.

Linacre, J. M. (1994). Many-facet Rasch measurement. Chicago, IL: MESA Press.

Masters, G. N. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149-174.

Mehrens, W. A. (1992). Using performance assessment for accountability purposes. Educational Measurement: Issues and Practice, 11 (1), 3-9, 20.

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. Educational Measurement: Issues and Practice, 14 (4), 5-8.

Rasch, G. (1960/1980). Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research, 1960. Expanded edition, Chicago: The University of Chicago Press, 1980.

Spandel, V. (1996). Seeing with New Eyes: A Guidebook on Teaching and Assessing Beginning Writers, 3rd ed. Portland, OR: Northwest Regional Educational Laboratory.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*: Rasch Measurement Chicago: MESA Press.

Wright, B. D., & Stone, M. H. (1979). *Best test design*: Rasch measurement. Chicago: MESA Press.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

TM035360

I. DOCUMENT IDENTIFICATION:

Title: The effect of benchmark selection on the assessed quality of student writing	
Author(s): Osborn Popp, S. E. & Ryan, J. M.	
Corporate Source:	Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☑

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature:	Printed Name/Position/Title: Sharon E. Osborn Popp, Faculty Associate/Research Associate, Arizona State University	
Organization/Address: 4531 W. Toldeo Street Chandler, AZ 85226	Telephone: (480) 705-0256	FAX:
	E-Mail Address: ssbornpo@asu.edu	Date: 11/07/2003

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC Clearinghouse on Assessment and Evaluation
University of Maryland, College Park
1129 Shriver Lab
College Park, MD 20742